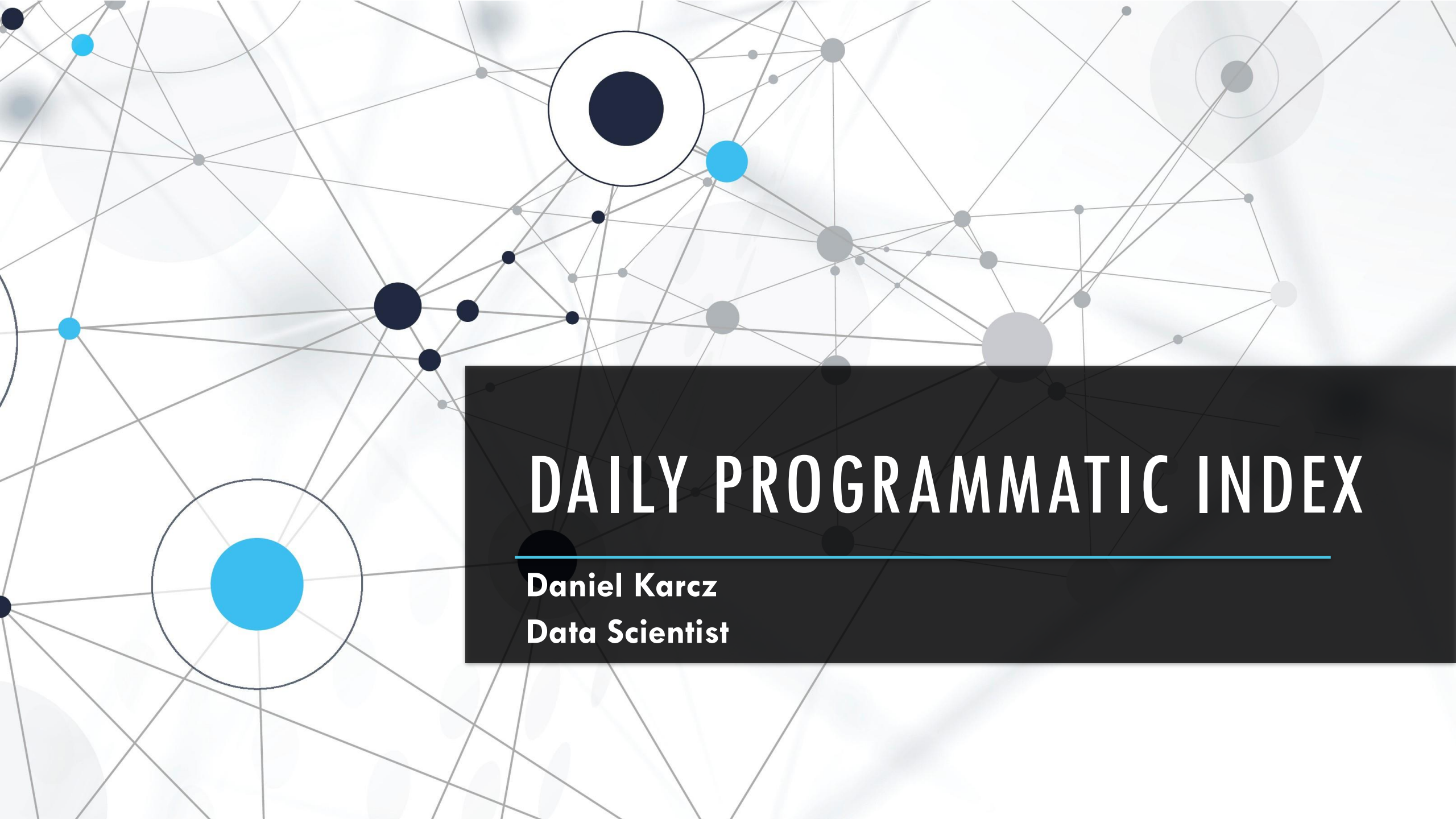


DAILY PROGRAMMATIC INDEX

Daniel Karcz

Data Scientist

CLOUD COMPUTATION PROJECT

DAILY PROGRAMMATIC INDEX

This presentation posits a hypothetical scenario between a Cost Estimator and Program Manager overseeing a cloud computation project at a company XYZ.

We will apply rigorous statistical methods on daily data, in order to provide proactive indications of cost over & under runs for a cloud computing project at corporation XYZ.

The proposed Daily Programmatic Index (DPI) value is an early warning system that will identify potentially significant cost swings.

MEASURING PROGRAMMATIC COST RISK VIA REAL TIME PULSES

Real time index pulses are a value add for cost estimators, risk managers, program managers, and earned value management practitioners.

Real time index pulses can be derived by:

1. Identification of programmatic cost drivers
2. Data dimensionality reduction via statistical techniques
3. Development of cost and risk indices
   - Overall supply chain index
   - Sector level supply chain indices
4. Holistic program management dashboards
   - Descriptive analytics - historical data
   - Predictive analytics - anticipated costs

WHY HIGH FREQUENCY DATA FOR COST ESTIMATION?

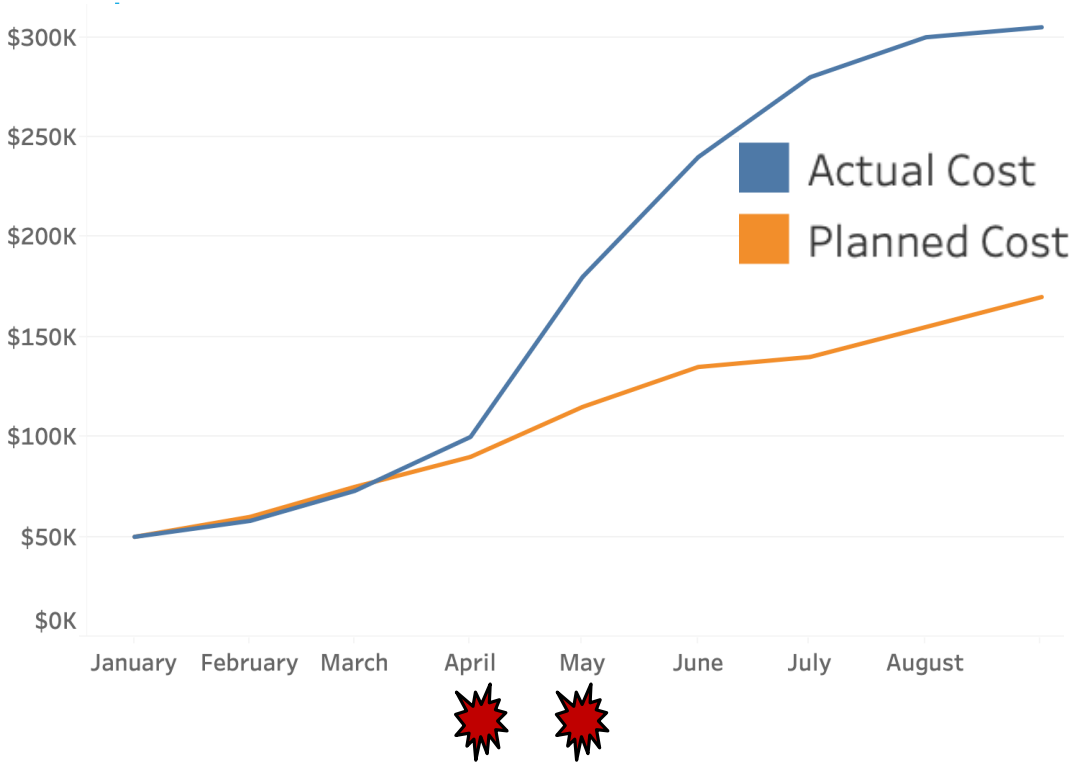
WHAT IS THE VALUE OF STATISTICAL METHODS?

1. Active monitoring of costs “vital signs”

2. Warning systems – cost problems

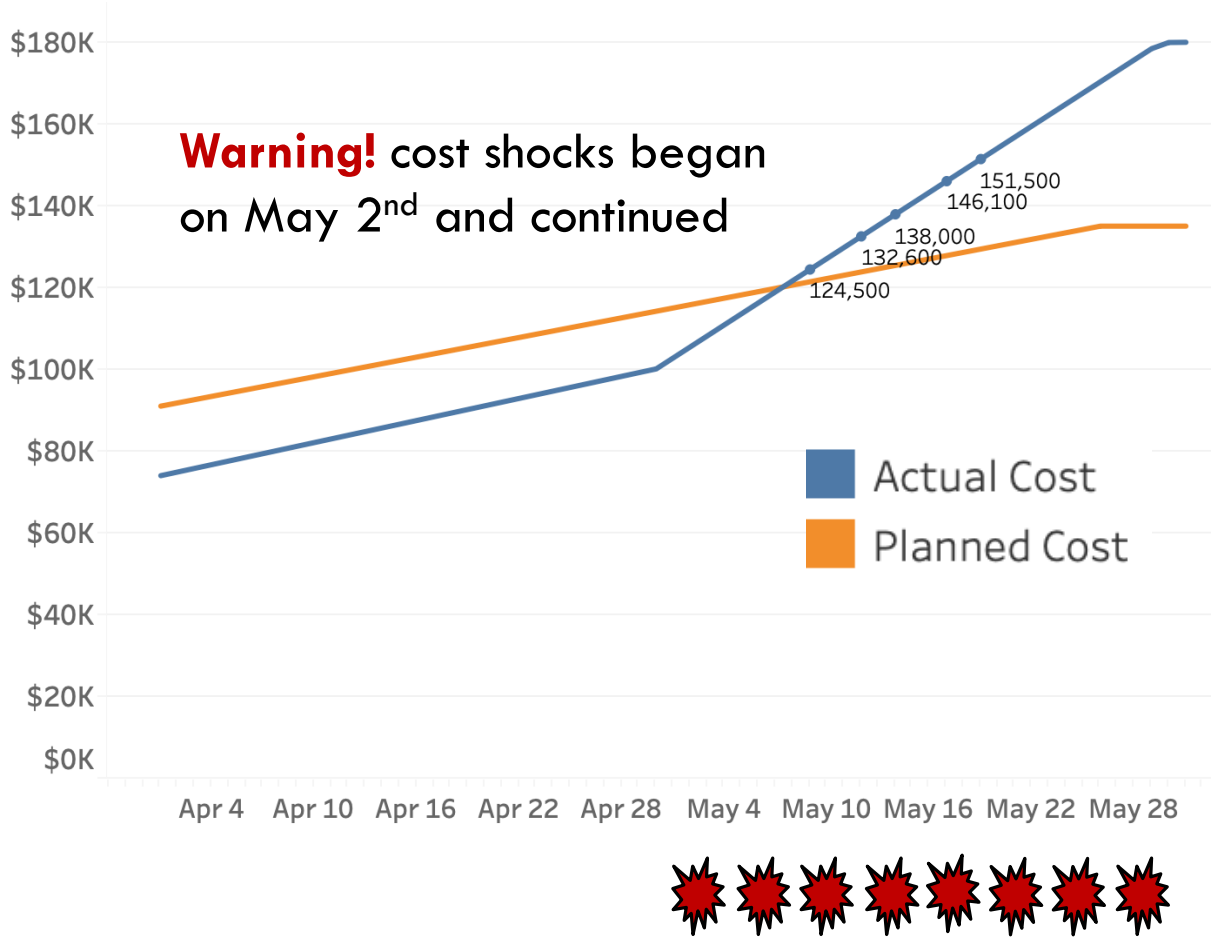
3. Data dimensionality reduction – more interpretable without loss of information

END OF MONTH COST REPORTS



- Monthly and quarterly reports
  - e.g. Contractor pay
  - e.g. Cloud computation usage
- Time consuming
- Time lag for identifying cost variances
- Not always standardized
- Subjective – error prone

DAILY PROGRAMMATIC INDEX (DPI) COST PULSE



DPI would have provided an early warning system. Program manager and cost estimator could have taken action to manage accelerating cloud costs which commenced on May 2nd. Instead of being surprised by end of month cost reports (i.e. **May 31st**).

INDICES CAN BE SPECIFIC OR HOLISTIC FOR THE SUPPLY CHAIN

An index can be built upon one of the supply chain stages:

1. Raw materials index (input ex: production rates of mined copper, steel, silver, timber, corn, etc.)
2. Manufacturer index (input ex: production rates of widgets, prices, etc.)
3. Distributer index (input ex: cost of fuel, cost of labor, delivery time, etc.)

Or an index can be based upon the entire supply chain of one industry. Example: gold industry.

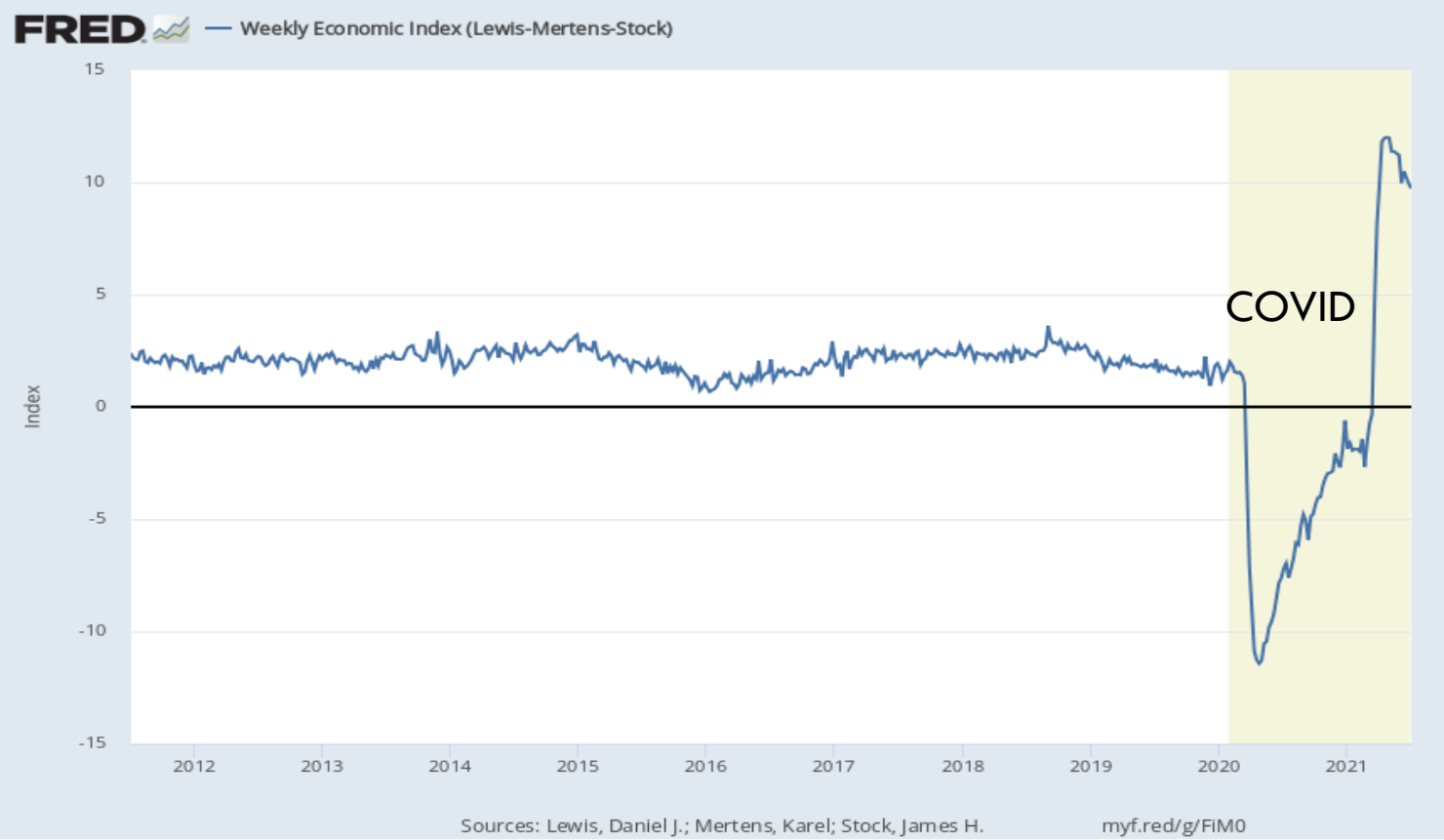
- Ore extracted from mine per week
- Grams of gold per ton of ore per week
- Cost of refining gold pre gram per week
- Cost of securely storing and shipping gold per week
- Spot market price of gold per week
- Retail jewelry customer purchase pattern behavior per week

Federal Reserve Economic Data (**FRED**) is a database maintained by the Research division of the Federal Reserve Bank of St. Louis that has more than 765,000 economic time series from 96 sources.

- Banking & Interest Rates & Reserves
- Consumer Price Index (CPI)
- Gross Domestic Product (GDP)
- Monetary Aggregates
- Producer Prices Indexes
- U.S. Trade and International Transactions

Inspiration

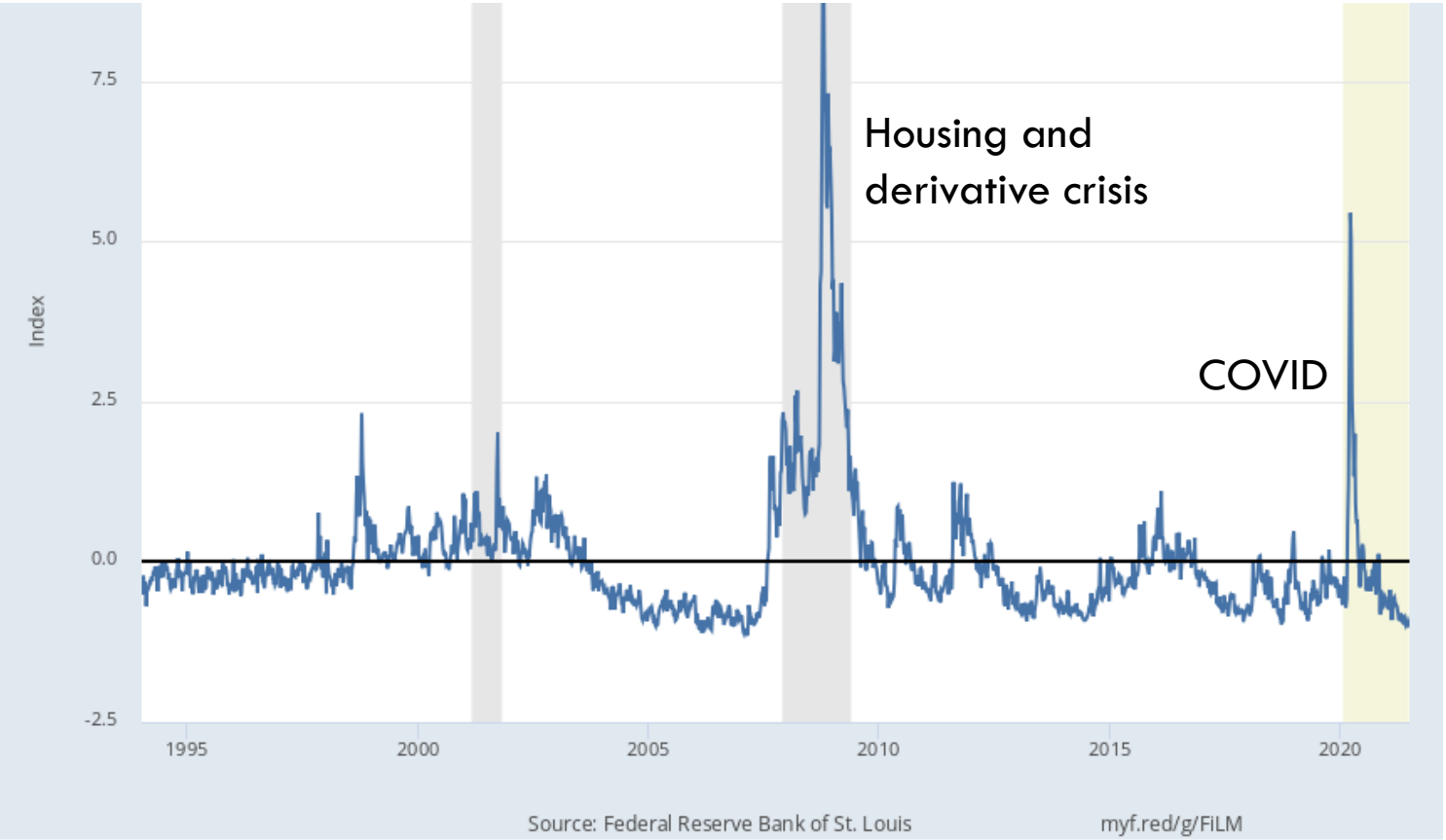
HOW CAN THESE INDEX METHODS BE APPLIED TO PROGRAM MANAGEMENT?



WEEKLY ECONOMIC INDEX: NEW YORK FED

The WEI is an index of real economic activity using timely and relevant high-frequency data. It represents the common component of ten different daily and weekly series covering consumer behavior, the labor market, and production.

1. Redbook same-store sales
2. Rasmussen Consumer Index
3. New claims for unemployment insurance
4. Continued unemployment insurance claims
5. Adjusted income/employment tax withholdings (from Booth Financial Consulting)
6. Railroad traffic originated (Association of American Railroads)
7. American Staffing Association Staffing Index
8. Steel production
9. Diesel, gasoline, and jet fuel
10. Average US electricity load



ST. LOUIS FED FINANCIAL STRESS INDEX

Measures the degree of financial stress in the markets and is constructed from 18 weekly data series, all of which are weekly averages of daily data series: seven interest rates, six yield spreads, and five other indicators. Each of these variables captures some aspect of financial stress.

How to Interpret the Index:
The average value of the index, which begins in late 1993, is designed to be zero. Thus, zero is viewed as representing normal financial market conditions. Values below zero suggest below-average financial market stress, while values above zero suggest above-average financial market stress.

DATASET FEATURES OF CORPORATE XYZ

Date	Type	Function	Unit	Price	Usage	Cost
2/3/2019	computation	active storage	gigabyte	$ 0.02	2982	$ 59.64
2/3/2019	computation	BigQuery Storage Write API	gigabyte	$ 0.03	3523	$ 88.08
2/3/2019	computation	Long-term storage	gigabyte	$ 0.01	37140	$ 371.40
2/3/2019	computation	Streaming inserts	megabyte	$ 0.01	8318	$ 83.18
2/3/2019	computation	BigQuery Storage Read API	terabyte	$ 1.10	3459	$3,804.90
2/3/2019	computation	queries	terabyte	$ 5.00	28.8	$ 144.00
2/3/2019	labor	Data Scientists	hourly rate	$ 57.30	20	$1,146.00
2/3/2019	labor	Database Administrators	hourly rate	$ 47.12	20	$ 942.40
2/3/2019	labor	Developers	hourly rate	$ 64.50	20	$1,290.00
2/3/2019	labor	Engineer	hourly rate	$ 95.80	20	$1,916.00
2/3/2019	labor	Data Scientists OT	overtime hourly rate	$ 85.95	0	$ -
2/3/2019	labor	Database Administrators OT	overtime hourly rate	$ 70.68	0	$ -
2/3/2019	labor	Developers OT	overtime hourly rate	$ 96.75	10	$ 967.50
2/3/2019	labor	Engineer OT	overtime hourly rate	$143.70	0	$ -
2/3/2019	Training	personal development session	rate per day per employee	$250.00	0	$ -
2/3/2019	utility	Electricity	KwH	$ 0.84	6940	$5,829.60
2/3/2019	utility	Space Rent	sq footage	$ 18.00	1	$ 18.00

TECHNICAL METHODS:

COST ESTIMATION WITH HIGH FREQUENCY DATA

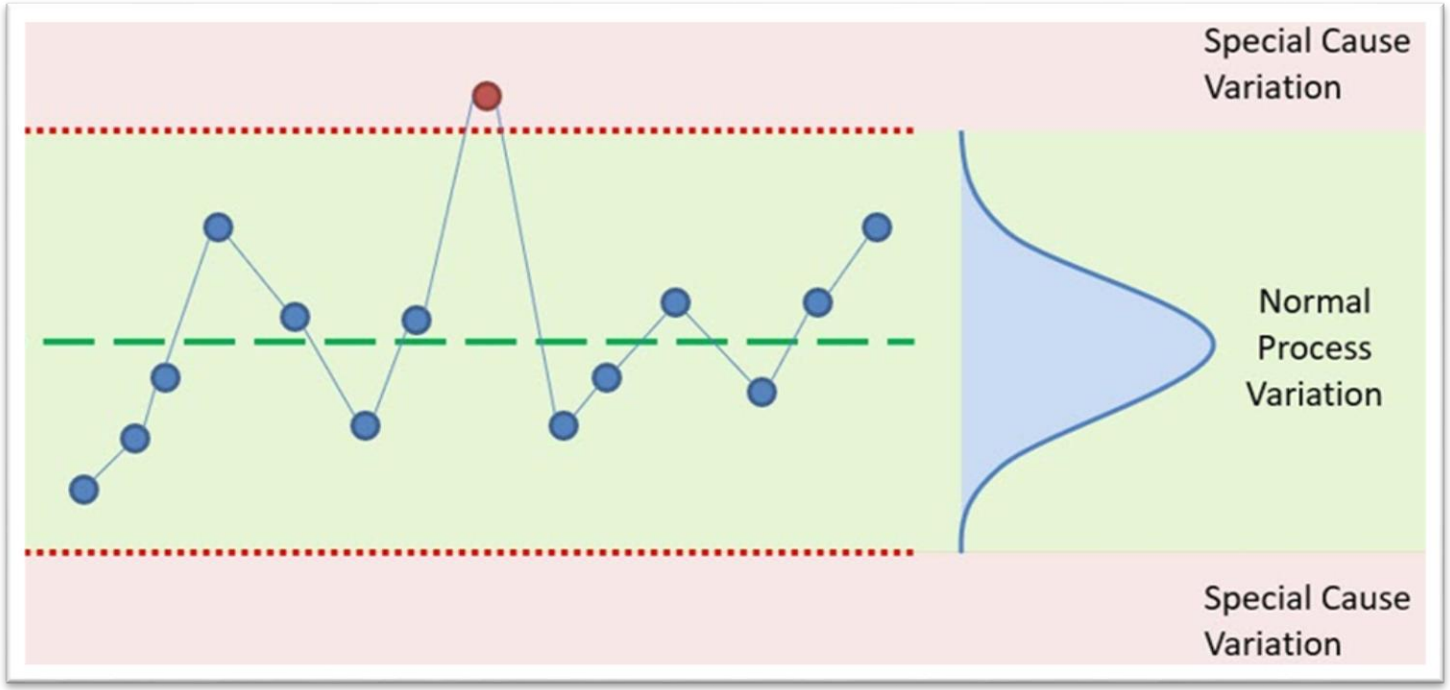
1. Control Charts

- Visualize whether data collected over time is stationary (i.e. “stable”) or if there are signals that could be out of the ordinary (i.e. “shocks”).

2. Principal Components Analysis (PCA)

- With increasing data availability data sets are often large and difficult to interpret. Principal component analysis (PCA) is a technique for reducing the dimensionality of such datasets, increasing interpretability while minimizing information loss.

METHOD 1: CONTROL CHARTS – VISUAL EXAMPLE



Outlier (i.e. “shocks”)

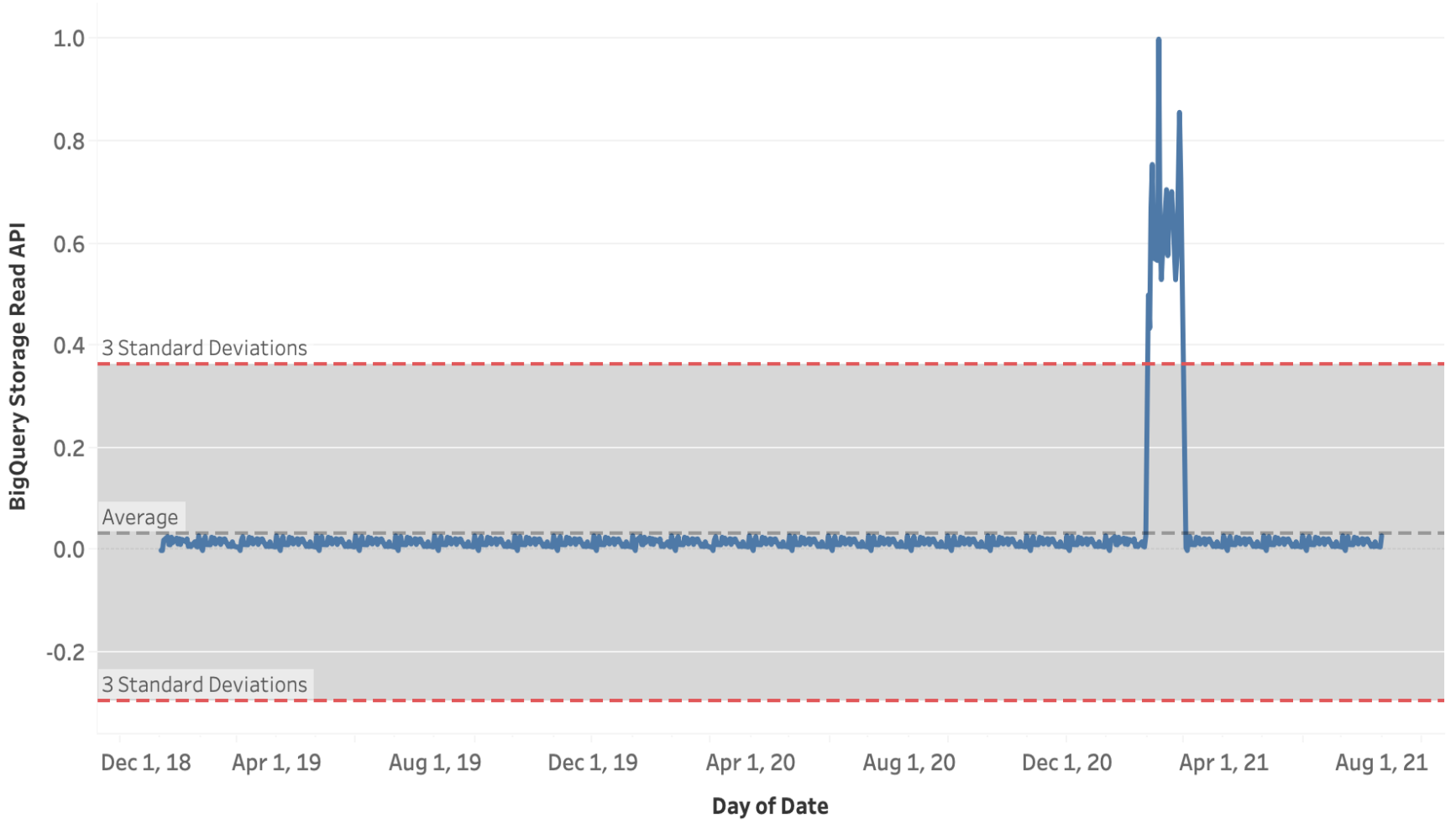
Average (i.e. “stable”)

CONTROL CHARTS

Query Storage Read API costs exploded in late February and early March due to subsidiary corporate XYZ and ABC mergers, and aggressive cloud testing.

Control charts would successfully alert a Program Manager or Cost Estimator that a specific cost is burning hot, and a cost valve must be turned to cool large deviances in cost.

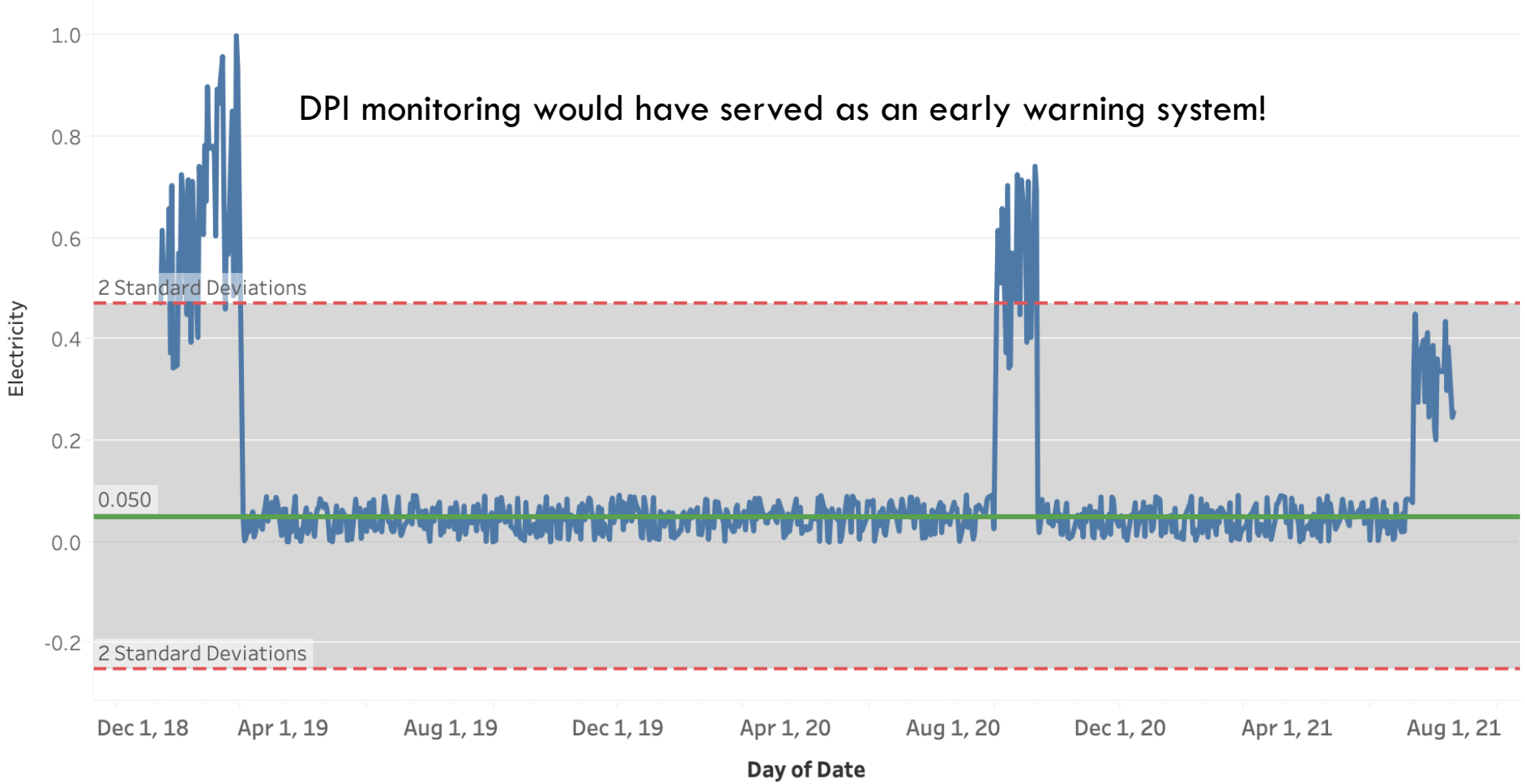
Notice the peak costs are above 3 standard deviations away from the average.



CONTROL CHARTS: ELECTRICITY & CLOUD COMPUTING

- The Average Electricity Rate in the U.S. is **10.42 cents per kilowatt-hour** (2021)

- **ENRON –** The 2000 California electricity crisis (**Western U.S. energy crisis**) was a situation in which the state of California had a shortage of electricity supply caused by market manipulations and capped retail electricity prices. Delays in approval of new power plants, and market manipulation decreased supply. This caused an **800% increase** in wholesale prices from April 2000 to December 2000.

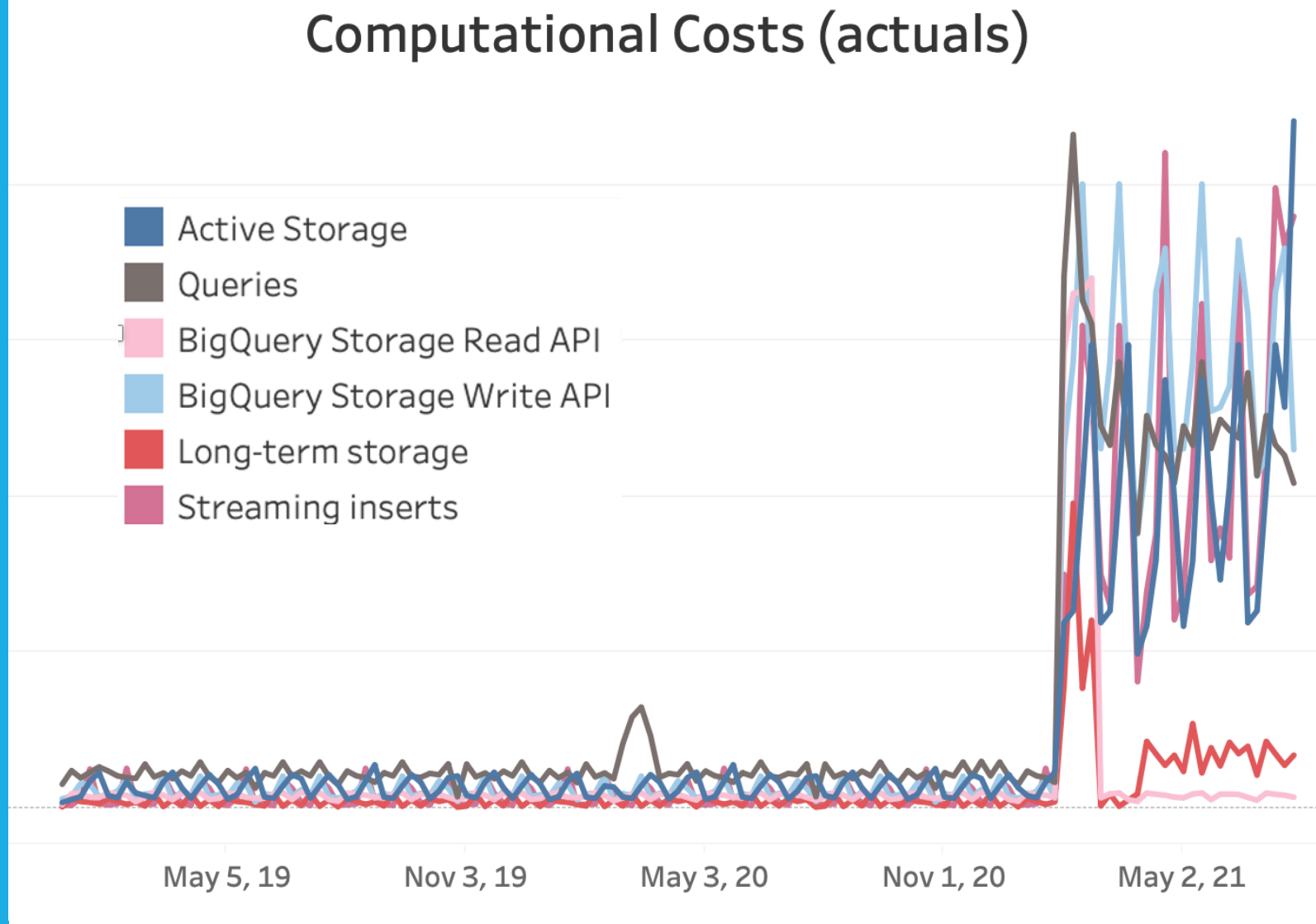


https://www.ferc.gov/industries-data/electric/general-information/addressing-2000-2001-western-energy-crisis

WHAT CAUSED THE COST SPIKES IN APRIL/MAY OF 2021?

WARNING SIGNS?

“XYZ” MERGED ALL SUBSIDIARY COMPANIES, WHICH INCREASED CLOUD USER BASE BY 7X.



METHOD 2: PRINCIPAL COMPONENTS VISUAL EXAMPLE

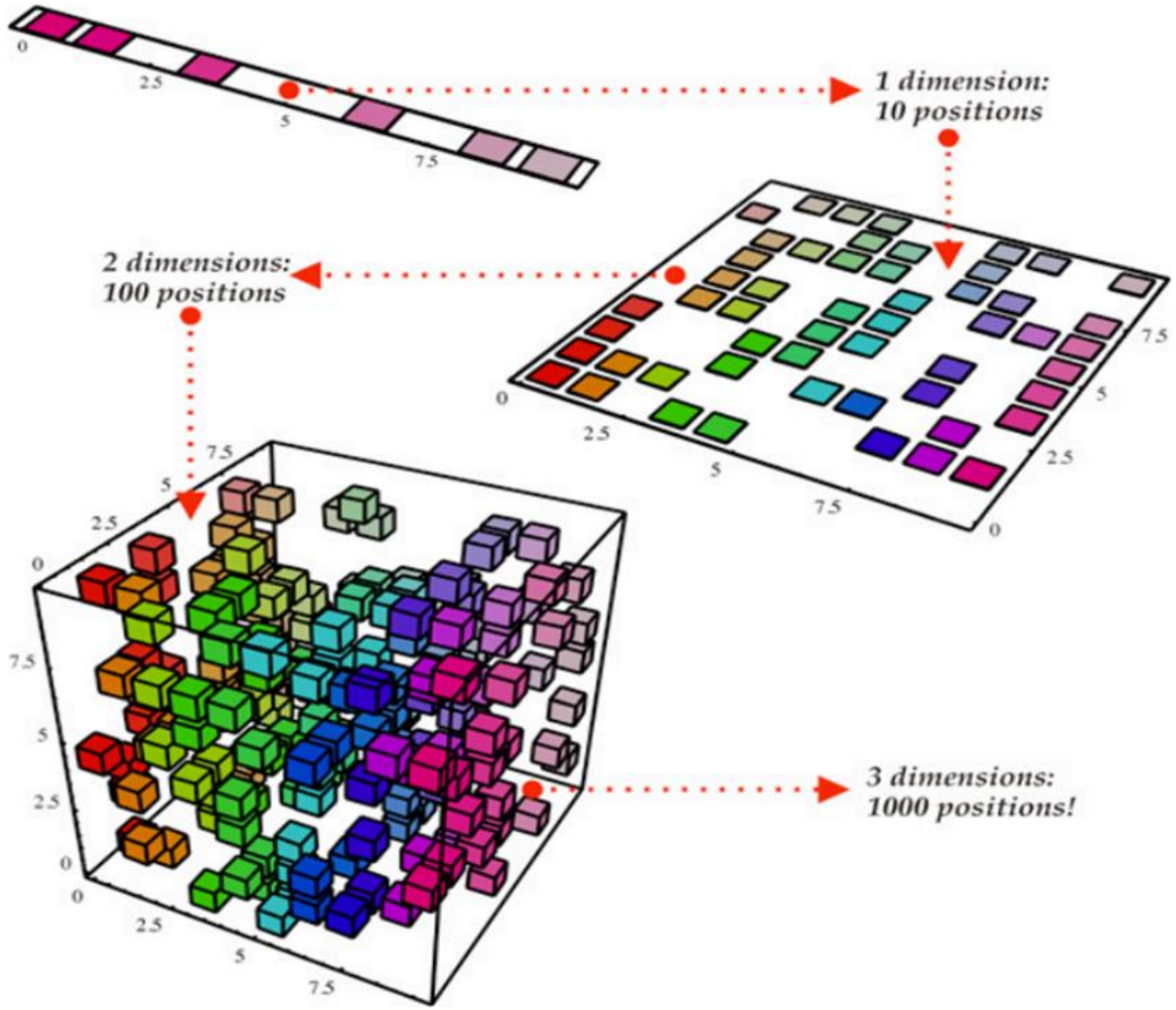


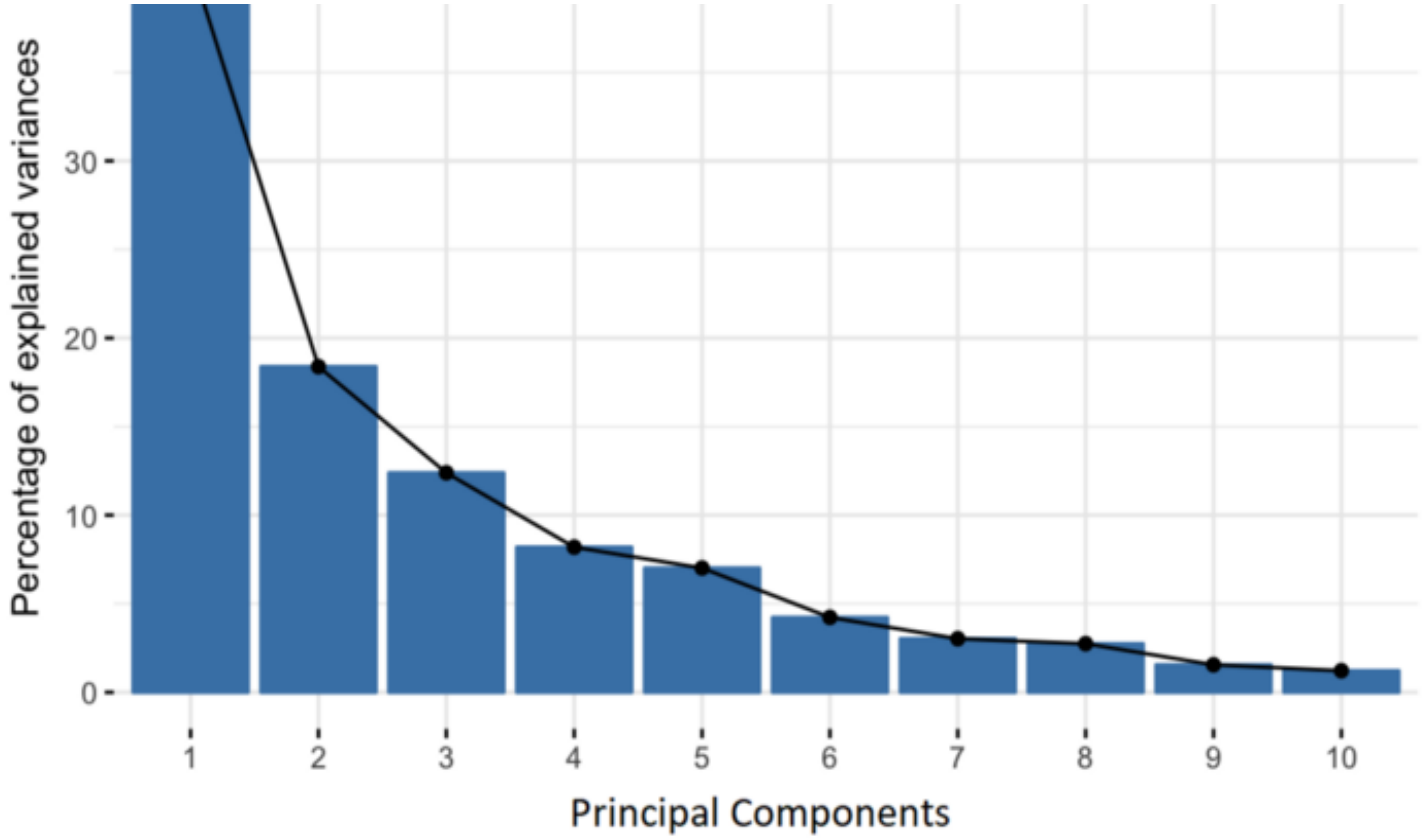
image source: http://www.turingfinance.com/artificial-intelligence-and-statistics-principal-component-analysis-and-self-organizing-maps/

Principal Components Analysis

Step 1: Standardize all data inputs to have same scale.

Step 2: Covariance Matrix Computation.

Step 3: Compute the eigenvectors and eigenvalues of the covariance matrix to identify the principal components.



PRINCIPAL COMPONENT ANALYSIS CONT.

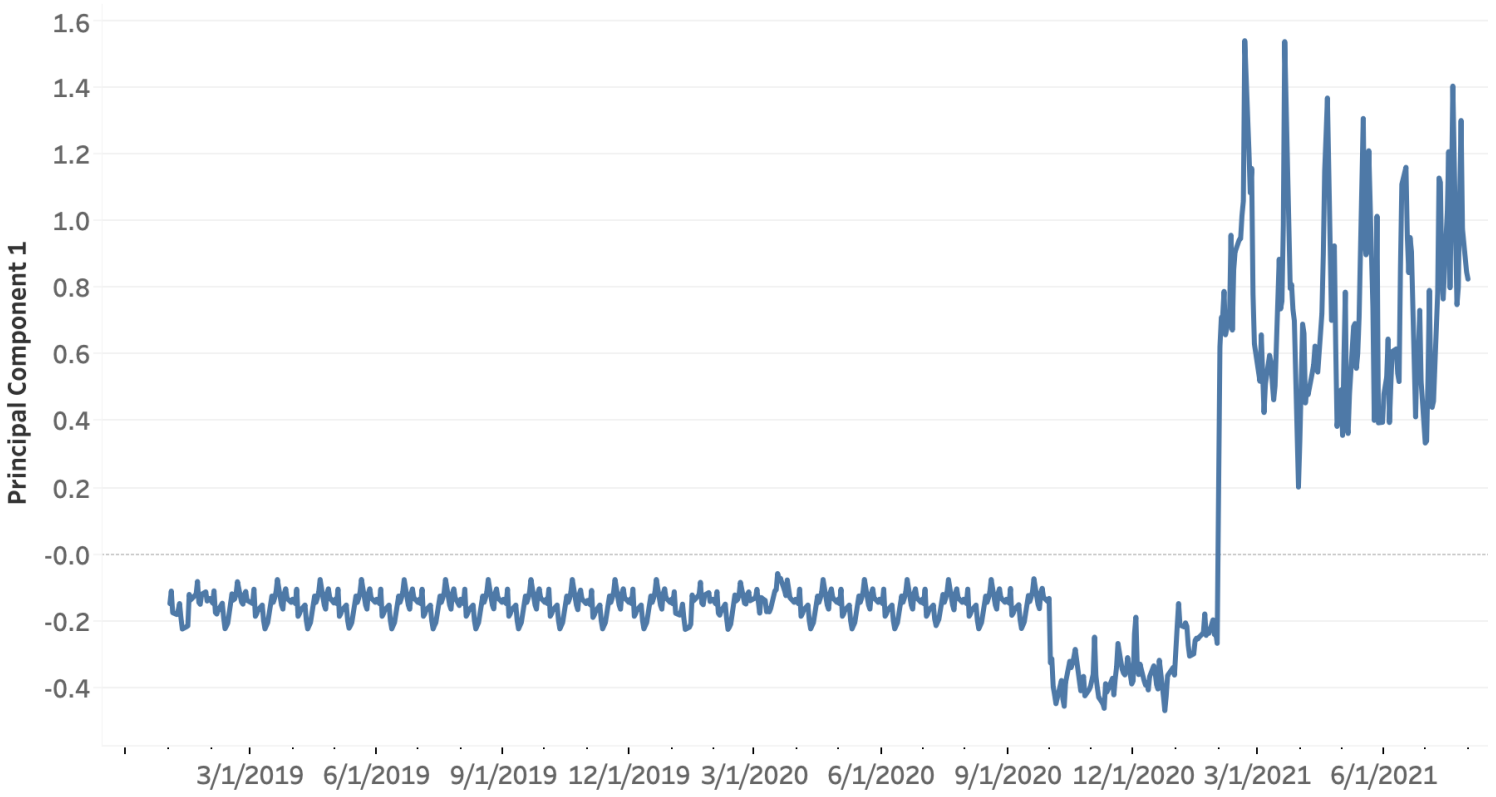
Principal Components Analysis

Step 3 continued: PCA puts maximum possible information in the first component, then maximum remaining information in the second and shown on (scree plot).

Trade off dimension reduction vs. interpretability

Organizing information in principal components this way, will allow you to reduce dimensionality without losing much information. Important! Principal components are less interpretable, when they are constructed as linear combinations of the initial variables.

Step 4: Recast the data long the Principal Component Axis.

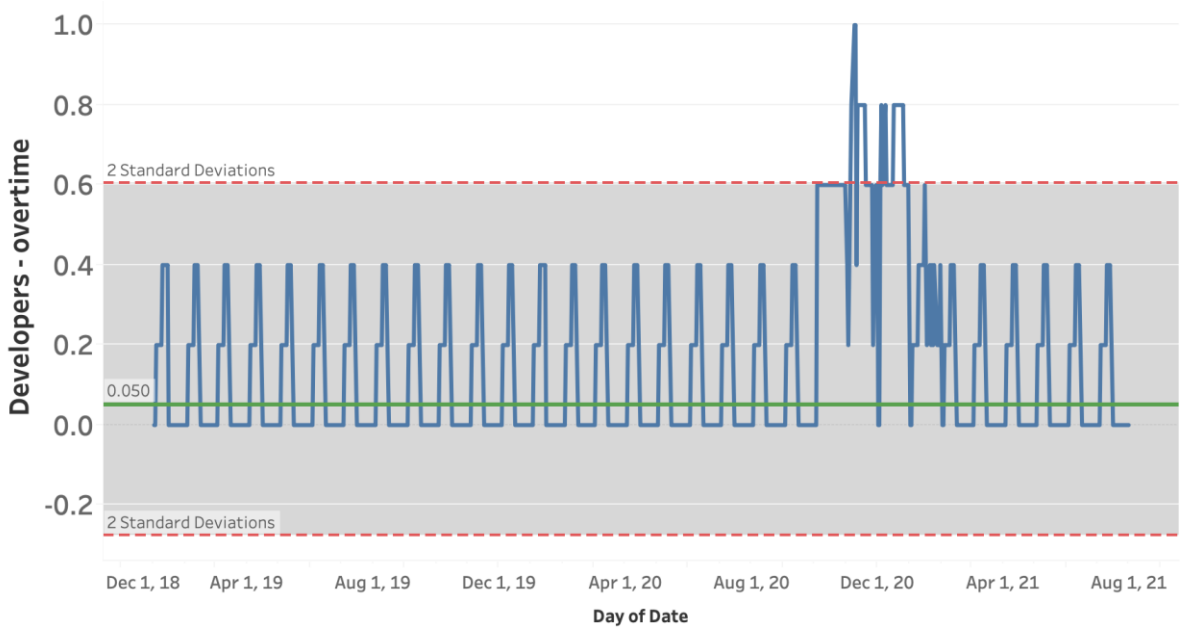


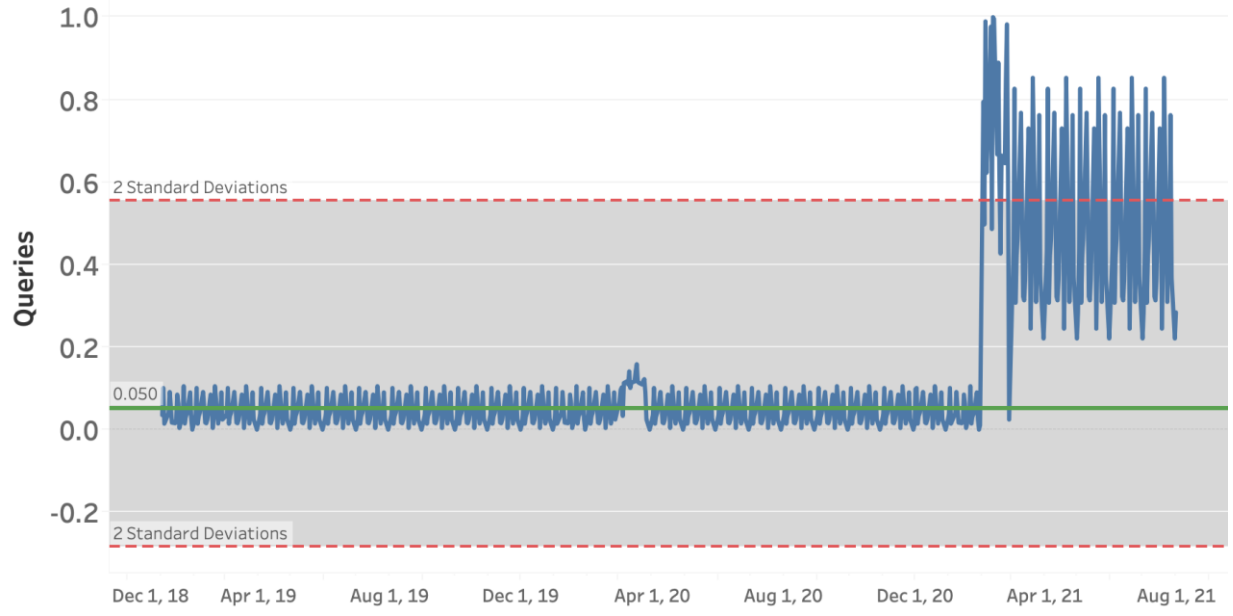
1ST PRINCIPAL COMPONENT

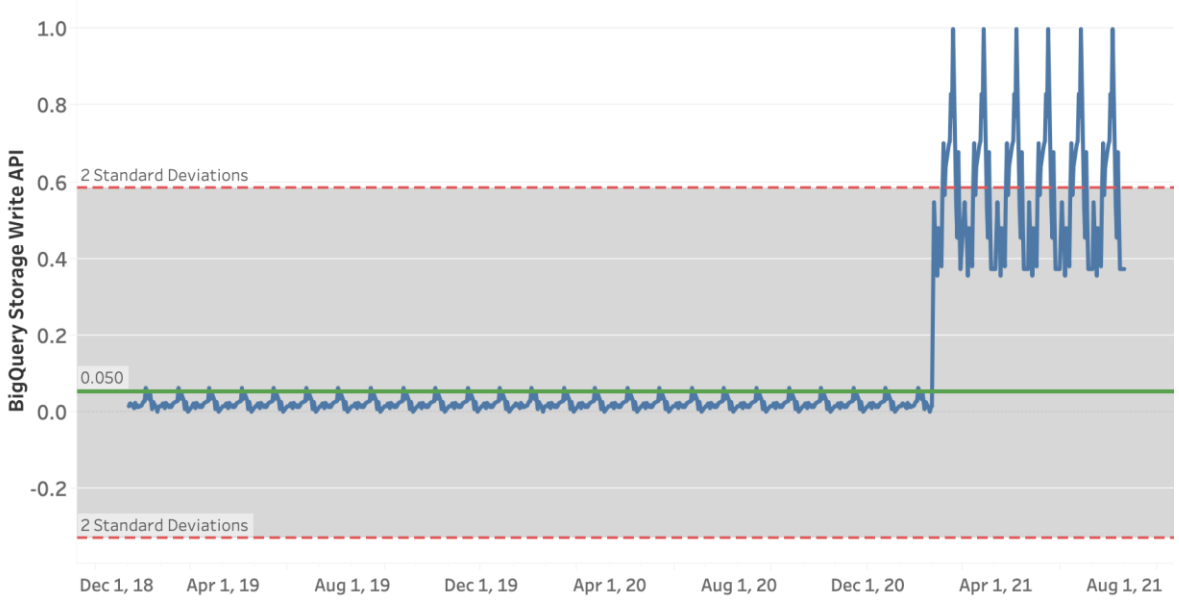
Step 1: Monitor DPI principal component

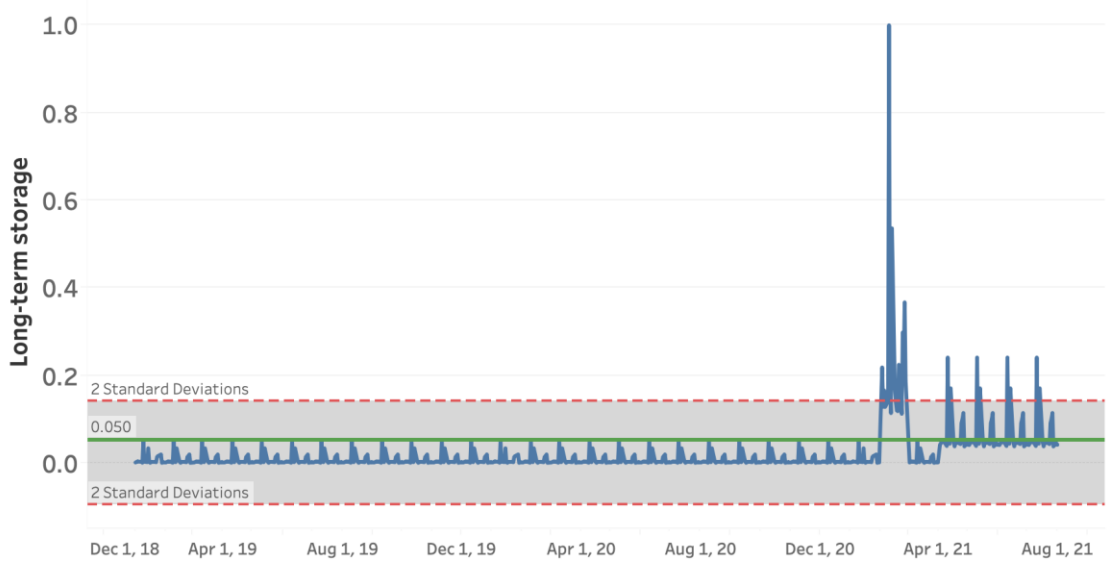
- Start from 40,000 foot view (PCA component 1 shock).
- Then observe each feature that was integrated into PCA. Did one cost driver change or a combination of cost drivers? Use control charts to observe cost shock deviations.

Step 2: if DPI identifies potentially significant cost swing during the day, then visualize control charts for each feature to identify cost driver's initiating swings.









Step 3: Program manager and cost estimator take proactive action to get ahead of budgetary contingency plans. Hypothetical examples below:

1) Overtime: employee biometric work sign in enabled, so overtime hours are tracked in source system of record at end of day.

✓ Action : temporarily suspend overtime pay until new cost estimation is developed and approved.

2) Cloud usage 7X increase

✓ Action: implement IT chargeback model to hold internal organizations accountable for computational costs.

3) Cloud computation increase

✓ Action: limited cloud usage per organization on a weekly basis.

Academic and Public Sources

Lewis, Daniel J, and James H Stock. “Weekly Economic Index (Lewis-Mertens-Stock).” *FRED*, Federal Reserve , 2 Sept. 2021, fred.stlouisfed.org/series/WEI.

Reid, Stuart. “Dimensionality Reduction Techniques.” *Turing Finance*, 22 June 2015, www.turingfinance.com/artificial-intelligence-and-statistics-principal-component-analysis-and-self-organizing-maps/.

“St. Louis FED Financial Stress Index.” *FRED*, Federal Reserve , 2 Sept. 2021, fred.stlouisfed.org/series/STLFSI2.